

WHAT IS CLAIMED IS:

1. In a data processing system, an apparatus comprising:

5 identity means for determining, for any of a plurality of data items in the system, a substantially unique identifier, said identifier depending on all of the data in the data item and only on the data in the data item; and

10 existence means for determining whether a particular data item is present in the system, by examining the identifiers of the plurality of data items.

2. An apparatus as in claim 1, further comprising:

15 local existence means for determining whether an instance of a particular data item is present at a particular location in the system, based on the identifier of the data item.

20 3. An apparatus as in claim 2, wherein each location contains a distinct plurality of data items, and wherein said local existence means determines whether a particular data item is present at a particular location in the system by examining the identifiers of the plurality of data items at said particular location in the system.

25 4. An apparatus as in claim 2, further comprising:

30 data associating means for making and maintaining, for a data item in the system, an association between the data item and the identifier of the data item; and

access means for accessing a particular data item using the identifier of the data item.

5. An apparatus as in claim 2, further comprising:

duplication means for copying a data item from a source to a destination in the data processing system, by providing said destination with the data item only if it is determined using the data identifier that the data item is not present at the destination.

6. An apparatus as in claim 4, further comprising:

assimilation means for assimilating a new data item into the system, said assimilation means invoking said identity means to determine the identifier of the new data item and invoking said data associating means to associate the new data item with its identifier.

7. An apparatus as in claim 4, further comprising:

duplication means for duplicating a data item from a source location to a destination location in the data processing system, based on the identifier of the data item, said duplication means invoking said local existence means to determine whether an instance of the data item is present at the destination location, and invoking said access means to provide said destination with the data item only if said local existence means determines that no instance of the data item is present at the destination.

8. An apparatus as in claim 7, further comprising:

backup means for making copies of data items in the system, said backup means maintaining a backup record of identifiers of data items backed up, and invoking duplication means to copy only those data items whose data identifiers are not recorded in the backup record.

9. An apparatus as in claim 8, further comprising:

5 recovery means for retrieving a data item previously backed up by said backup means, based on the identifier of the data item, said recovery means using the backup record to identify the data item, and invoking access means to retrieve the data item.

10 10. An apparatus as in claim 2, wherein a location is a computer among a network of computers, the apparatus further comprising:

15 remote existence means for determining whether a data item is present at a remote location in the system from a current location in the system, based on the identifier of the data item, said remote location using local existence means at the remote location to determine whether the data item is present at the remote location, and providing the current location with an indication of the presence of the data item at the remote location.

20 11. An apparatus as in claim 4, wherein a location is a computer among a network of computers, the apparatus further comprising:

25 requesting means for requesting a data item at a current location in the system from a remote location in the system, based on the identifier of the data item, said remote location using access means at the remote location to obtain the data item and to send it to the current location if it is present.

12. An apparatus as in claim 1, further comprising:

30 context means for making and maintaining a context association between at least one contextual name of a data item in the system and the identifier of the data item; and

referencing means for obtaining the identifier of a data item in the system given a contextual name for the data item, using said context association.

5 13. An apparatus as in claim 12, further comprising:

assignment means for assigning a data item to a contextual name, invoking said identity means to determine the identifier of the data item, and invoking said context means to make or modify the context association between the contextual name of the data item and the identifier of the data item.

14. An apparatus as in claim 12, further comprising:

15 data associating means for making and maintaining, for a data item in the system, an association between the data item and the identifier of the data item;

20 access means for accessing a particular data item using the identifier of the particular data item; and

contextual name access means for accessing a data item in the system for a given context name of the data item, determining the data identifier associated with the given context name, and invoking said access means to access the data item using the data identifier.

15. An apparatus as in claim 11, further comprising:

30 transparent access means for accessing a data item from one of several locations, using the identifier of the data item, said transparent access means invoking said local existence means to determine if the particular data item is present at the current location, and, in the case when the particular data item is not present at the

current location, invoking said requesting means to obtain the data item from a remote location.

16. An apparatus as in claim 15, further comprising:

5 identifier copy means for copying an identifier of a data item from a source location to a destination location.

17. An apparatus as in claim 15, further comprising:

10 context means for making and maintaining a context association between a contextual name of a data item in the system and the identifier of the data item;  
context copy means for copying a data item from a source location to a destination location, given the  
15 contextual name of the data item, by copying only the context association between the contextual identifier and the data identifier from the source location to the destination location; and

transparent referencing means for obtaining a  
20 data item from one of several locations the system given a contextual name for the data item, said transparent referencing means invoking said context association to determine the data identifier of a data item given a contextual name, and invoking said transparent access  
25 means to access the data item from one of several locations given the identifier of the data item.

18. An apparatus as in claim 1, wherein at least some of said data items are compound data items, each compound data item including at least some component  
30 data items in a fixed sequence, and wherein the identity means determines the identifier of a compound data item based on each component data item of the compound data item.

19. An apparatus as in claim 18, wherein said compound data items are files and said component data items are segments, and wherein the identity means determines the identifier of a file based on the identifier of each data segment of the file.

20. An apparatus as in claim 18, wherein said compound data items are directories and said component data items are files or subordinate directories, and wherein the identity means determines the identifier of a given directory based on each file and subordinate directory within the given directory.

21. An apparatus as in claim 11, further comprising:  
means for advertising a data item from a location in the system to at least one other location in the system, said means for advertising providing each of said at least one other location with the data identifier of the data item, and providing the data item to only those locations of said other locations that request said data item in response to said providing.

22. An apparatus as in claim 18, further comprising:  
local existence means for determining whether a particular data item is present at a particular location in the system, based on the identifier of the data item; and

compound copy means for copying a data item from a source to a destination in the data processing system, said compound copy means invoking said local existence means to determine whether the data item is present at the destination, and to determine, when the data item is a compound data item, whether the component data items of the compound data item are present at the destination, and providing said destination with the data

item only if said local existence means determines that  
the data item is not present at the destination, and  
providing said destination with each component data item  
only if said local existence means determines that the  
5 component data item is not present at the destination.

23. An apparatus as in claim 11, further  
comprising:

means for verifying the integrity a data item  
obtained from said requesting means in response to  
10 providing said requesting with a particular data  
identifier, to confirm that the data item obtained from  
the requesting means is the same data item as the data  
item requested, said verifying means invoking said  
identity means to determine the data identifier of the  
15 obtained data item, and comparing said determined data  
identifier with said particular data identifier to verify  
said obtained data item.

24. An apparatus as in claim 2, wherein a  
location is at least one of a storage location and a  
20 processing location, and wherein a storage location is at  
least one of a data storage device and a data storage  
volume, and wherein a processing location is at least one  
of a data processor and a computer.

25. An apparatus as in claim 3, wherein at  
25 least some of said data items are compound data items,  
each compound data item including at least some component  
data items in a fixed sequence, and wherein the identity  
means determines the identifier of a compound data item  
based on the identifier of each component data item of  
30 the compound data item.

26. An apparatus as in claim 3, further  
comprising:

context associating means for making and  
maintaining a context association, for any data item in  
the system, between the identifier of the data item and  
at least one contextual name of the data item at a  
5 particular location in the system;

means for obtaining the identifier of a data  
item in the system given a contextual name for the data  
item at a particular location in the system; and

logical copy means for associating the data  
10 identifier corresponding to a contextual name at a source  
location with a contextual name at a destination location  
in the data processing system.

27. An apparatus as in claim 25, wherein said  
compound data items are files and said component data  
15 items are segments, and wherein the identity means  
determines the identifier of a file based on the  
identifier of each data segment of the file.

28. An apparatus as in claim 25, further  
comprising:  
20 compound copy means for copying a data item  
from a source location to a destination location in the  
data processing system, said compound copy means invoking  
said local existence means to determine whether the data  
item is present at the destination, and to determine,  
25 when the data item is a compound data item, whether the  
component data items of the compound data item are  
present at the destination, and providing said  
destination with the data item only if said local  
existence means determines that the data item is not  
30 present at the destination, and providing said  
destination with each component data item only if said  
local existence means determines that the component data  
item is not present at the destination.





determining, using said data identifier,  
whether said data item is present at said destination  
location; and

5 based on said determining, providing said  
destination location with said data item only if said  
data item is not present at said destination.

34. A method as in claim 33, wherein said  
given data item is a compound data item having a  
plurality of component data items, the method further  
10 comprising the steps of:

for each data item of said component data  
items,

obtaining the component data  
15 identifier of the data item by determining a  
substantially unique identifier for the data  
item, said identifier depending on all of the  
data in the data item and only on the data in  
the data item;

determining, using said obtained  
20 component data identifier, whether said data  
item is present at said destination; and

based on said determining, providing  
said destination with said data item only if  
said data item is not present at said  
25 destination.

35. A method for determining whether a  
particular data item is present in a data processing  
system, the method comprising the steps of:

(A) for each data item of a plurality of data  
30 items in the system,

(i) determining a substantially unique  
identifier for the data item, said identifier  
depending on all of the data in the data item  
and only on the data in the data item; and

(ii) making and maintaining a set of  
identifiers of said plurality of data items;  
and

(B) for the particular data item,

(i) determining a particular  
substantially unique identifier for the data  
item, said identifier depending on all of the  
data in the data item and only on the data in  
the data item; and

(ii) determining whether said particular  
identifier is in said set of data items.

36. A method of backing up, of a plurality of  
data items, data items modified since a previous backup  
time in a data processing system, the method comprising  
the steps of:

(A) maintaining a backup record of identifiers  
of data items backed up at the previous backup  
time; and

(B) for each of said plurality of data items,  
(i) determining a substantially unique  
identifier for the data item, said  
identifier depending on all of the data in  
the data item and only on the data in the  
data item;

(ii) determining those data items of the  
plurality of data items whose identifiers  
are not in the backup record; and

(iii) based on said determining, copying  
only those data items whose data  
identities are not recorded in the backup  
record.

37. A method as in claim 36, further  
comprising the step of:

recording in the backup record the identifiers  
of those data items copied in said step of copying.

38. A method of locating a particular data  
item at a location in a data processing system, the  
5 method comprising the steps of:

- (A) determining a substantially unique  
identifier for the data item, said identifier  
depending on all of the data in the data item  
and only on the data in the data item;
- 10 (B) requesting the particular data item by  
sending the data identifier of the data item  
from the requestor location to at least one  
location of a plurality of provider locations  
in the system; and
- 15 (C) on at least some of said provider  
locations,
  - (a) for each data item of a plurality of  
data items at said provider locations,
    - 20 (i) determining a substantially unique  
identifier for the data item, said  
identifier depending on all of the data in  
the data item and only on the data in the  
data item; and
    - (ii) making and maintaining a set of  
25 identifiers of data items,
  - (b) determining, based on said set of  
identifiers, whether the data item  
corresponding to the requested data  
identifier is present at said provider  
30 location; and
  - (c) based on said determining, when said  
provider location determines that the  
particular data item is present at the  
provider location, notifying said  
35 requestor that the provider has a copy of  
the given data item.

39. The method of claim 38, further comprising the steps of:

- 5 (a) for each data item of a plurality of data items at said provider locations,  
making and maintaining an association between the data item and the identifier of the data item,  
10 (b) in response to said notifying, said client location copying said data item from one of said responding remote locations, using said association to access the data item given the data identifier.

40. A method of locating a particular data item among a plurality of locations, each of said locations having a plurality of data items, the method comprising the steps of:

- 15 determining, for the particular data item and for each data item of the plurality of data items, a substantially unique identifier for the data item, said identifier depending on all of the data in the data item and only on the data in the data item; and  
20 determining the presence of the particular data item in each of said plurality of locations by determining whether the identifier of the particular data item is present at each of said locations.  
25

41. The method of claim 30, wherein said step of accessing further comprises the steps of, for a given data identifier and for a given current location and a remote location in the system:

- 30 determining whether the data item corresponding to the given data identifier is present at the current location, and  
based on said determining, if said data item is not present at the current location, fetching the data

item from a remote location in the system to the current location.

42. The method of claim 41, further comprising the steps of:

5           for each contextual name at a location,  
            making and maintaining a context  
association between the context name of a data item and  
the identifier of said data item, and when some context  
association changes at said current location, and  
10           notifying said remote location of a  
modification to the context association.

43. The method of claim 42, further comprising the step of:

15           at said remote location, updating the  
association between the contextual identifier of the data  
item and the identifier of the data item.

44. The method of claim 43, further comprising the step of:

20           from said remote location, notifying all other  
locations that said data item has been modified, by  
providing the contextual identifier and data identifier  
of said data item to said other locations.

45. The method of claim 44, further comprising the step of, at each location notified that the data item  
25           has been modified:

            modifying an association between the contextual  
identifier of the data item and the data identifier of  
the data item, to record that the data item has been  
modified.

30           46. A method of eliminating a data item at a  
given location in a data processing system when said data

item can be obtained from another location in the system,  
the method comprising the steps of:

5           determining a substantially unique identifier  
for the data, said identifier depending on all of the  
data in the data item and only on the data in the data  
item;

          making and maintaining a source association  
between the data identifier and at least one location at  
which said data item is known to be present; and  
10           based on said source association, if said data  
item is present at said other location, removing the data  
item from the given location.

47. A method of deleting a data item from a  
location in a data processing system, the method  
15           comprising the steps of:

          for each of a plurality of data items in the  
system:

          determining a substantially unique identifier  
for the data, said identifier depending on all of the  
20           data in the data item and only on the data in the data  
item; and

          making and maintaining, an association between  
each of the data items and the unique identifier of the  
data items; and

25           for a given data item:

          determining a substantially unique identifier  
for the data, said identifier depending on all of the  
data in the data item and only on the data in the data  
item; and

30           determining whether a contextual identifier or  
a compound data item or a remote processor in the system  
refers to the unique identifier of the data item, and  
based on said determining, deleting said data item and  
its association if no other contextual identifier or  
35           compound data item or remote processor refers to said  
data item.

48. The method of claim 47, wherein said determining is based on a use count for the data item, and wherein said data item is deleted only if said use count indicates that no other contextual identifier or compound data item or remote processor in the system refers to the data item.

49. A method of substantially synchronizing data items at a client location in a data processing system after a period of independent changes on the client and another location in the system, given a context, the method comprising the steps of:

making and maintaining a list of changes to the context association between each context name of a data item and the identifier of said data item, in the given context and during the period of independent change;

obtaining the list of changes from the other location for the given context; and,

for each context name in the list of changes  
updating the context identifier  
20 associations at the client whenever it is determined that  
the context association of the given context name changed  
either only at the client or only at the other location  
during the period if independent changes; and

performing a conflict-resolution task such  
25 as notifying an operator of the client location, whenever  
it is determined that the context association changed at  
both the client and the other location.

50. A method as in claim 49, wherein said  
lists are maintained as queues based on a temporal order,  
30 and wherein, at said client location, said replacing is  
based on said temporal order.

51. A method of maintaining at least a predetermined number of copies of a given data item in a data processing system, at different locations in the



data processing system, said data processing system being one wherein data is identified by a substantially unique identifier, said identifier depending on all of the data in the data item and only on the data in the data item, and wherein any data item in the system may be accessed using only the identifier of the data item, the method comprising the steps of:

(i) sending, from a first location in the system, the data identifier of the given data item to other locations in the system; and

(ii) in response to said sending, at each of said other locations,

(A) determining whether the data item corresponding to the data identifier is present at the other location, and based on said determining, and

(B) informing said first location whether said data item is present at the other location; and

(iii) in response to said informing from said other locations, at said first location,

(A) determining whether said data item is present in at least the predetermined number of other locations, and based on said determining,

(B) when less than the predetermined number of other locations have a copy of the data item, requesting some locations that do not have a copy of the data item make a copy of the data item.

52. A method as in claim 51, wherein said step (iii) further comprises the step of:

(C) when more than the predetermined number of other locations have a copy of the data item, requesting some locations that do have a copy of the data item delete the copy of the data item.

53. A method as in any of claims 30-52, wherein said data items are at least one of a file, a database record, a message, a data segment, a data block, a directory, and an instance of an object class.